Optimal Decision Feedback Equalizer for M-PAM Signals Using a Support Vector Machine Solution

S. Chen, A.K. Samingan and L. Hanzo

Indexing terms: Decision feedback equalizer, symbol error rate, support vector machines

We propose a method of designing the linear-combiner decision feedback equalizer (DFE) for M-PAM signals using a support vector machine (SVM) technique. This SVM design achieves asymptotically the minimum symbol error rate (MSER) solution and can be computed efficiently.

Introduction

The linear-combiner DFE is widely used in practice, as it provides a good trade off between performance and complexity. The most popular design strategy is the minimum mean square error (MMSE) design with an advantage of simple adaptive implementation. A better design in terms of symbol error rate (SER) is to choose the equalizer coefficients to minimize SER directly [1]. However, unlike the quadratic mean square error surface, the SER surface can be highly irregular and a gradient algorithm cannot generally guarantee to converge to a global minimum. The idea of SVMs [2] originates from finding an optimal hyperplane to separate two classes with maximum margin, and this is very relevant to equalization. The method of SVM has been applied to design the DFE for binary signals [3]. This SVM solution is easy to compute and is asymptotically the MSER solution. We extend the results of [3] to the general case of M-PAM symbols.

The DFE structure

We will assume that the real-valued channel generates the received signal samples of:

$$y(k) = \sum_{i=0}^{n_a-1} a_i s(k-i) + e(k) \tag{1}$$

where n_a is the channel impulse response (CIR) length, a_i denotes the channel taps, the Gaussian white noise $e(k)$ has zero mean and variance σ_e^2, and the M-PAM symbol $s(k)$ takes values from the set $\{s_i = 2i - M - 1, 1 \le i \le M\}$. The linear-combiner DFE produces an estimate $\hat{s}(k-d)$ of $s(k-d)$ by quantizing the filter output of:

$$f(\mathbf{y}(k), \hat{\mathbf{s}}_b(k)) = \mathbf{w}^T \mathbf{y}(k) + \mathbf{b}^T \hat{\mathbf{s}}_b(k) \tag{2}$$

where $\mathbf{y}(k) = [y(k)\ y(k-1)\cdots y(k-m+1)]^T$ and $\hat{\mathbf{s}}_b(k) = [\hat{s}(k-d-1)\cdots\hat{s}(k-d-n)]^T$ are the observation and past detected symbol vectors, respectively, while $\mathbf{w} = [w_0\ w_1\cdots w_{m-1}]^T$ and $\mathbf{b} = [b_1\cdots b_n]^T$ are the coefficient vectors of the feedforward and feedback filters, respectively. Without the loss of generality, $d = n_a - 1$, $m = n_a$ and $n = n_a - 1$ are chosen, as this choice is sufficient to guarantee the linear separability [1].

Define $\mathbf{s}_f(k) = [s(k)\cdots s(k-d)]^T$, $\mathbf{s}_b(k) = [s(k-d-1)\cdots s(k-d-n)]^T$, and

$$F_1 = \begin{bmatrix} a_0 & a_1 & \cdots & a_{n_a-1} \\ 0 & a_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_1 \\ 0 & \cdots & 0 & a_0 \end{bmatrix} \tag{3}$$

$$F_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ a_{n_a-1} & 0 & \ddots & \vdots \\ a_{n_a-2} & a_{n_a-1} & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ a_1 & \cdots & a_{n_a-2} & a_{n_a-1} \end{bmatrix} \tag{4}$$

the $m \times (d+1)$ and $m \times n$ CIR matrices, respectively. Under the assumption of correct decision feedback, we have $\hat{\mathbf{s}}_b(k) = \mathbf{s}_b(k)$, and the decision feedback translates the original signal space $\mathbf{y}(k)$ into a new space $\mathbf{r}(k)$ [1]:

$$\mathbf{r}(k) \triangleq \mathbf{y}(k) - F_2\ \hat{\mathbf{s}}_b(k) \tag{5}$$

In the translated space, the DFE can be described by:

$$f(\mathbf{r}(k)) = \mathbf{w}^T\mathbf{r}(k) \tag{6}$$

Let the $N_f = M^{d+1}$ possible sequences of $\mathbf{s}_f(k)$ be $\mathbf{s}_{fj}$, $1 \le j \le N_f$. The set of the noiseless channel states in the translated space, namely

$$R \triangleq \{\mathbf{r}_j = F_1\ \mathbf{s}_{fj},\ 1 \le j \le N_f\} \tag{7}$$

can be partitioned into the M subsets:

$$R^{(i)} \triangleq \{\mathbf{r}_j \in R : s(k-d) = s_i\}, \quad 1 \le i \le M \tag{8}$$

Lemma 1 $R^{(i+1)}$ is a translation of $R^{(i)}$, and $R^{(l)}$, $1 \le l \le M$, are linearly separable.

Proof: From the definitions of $R^{(i)}$ and F_1,

$$R^{(i+1)} = R^{(i)} + (s_{i+1} - s_i)\mathbf{a}_{\text{rev}} = R^{(i)} + 2\mathbf{a}_{\text{rev}} \tag{9}$$

where $\mathbf{a}_{\text{rev}} = [a_{n_a-1} \cdots a_1 a_0]^T$. To prove the linear separability, consider the hyperplane

$$H_i(\mathbf{r}) \triangleq \hat{\mathbf{w}}^T \mathbf{r} + 2\left(\frac{M}{2} - i\right) \hat{\mathbf{w}}^T \mathbf{a}_{\text{rev}} = 0 \tag{10}$$

with

$$\hat{\mathbf{w}} = \left[0\ 0 \cdots 0\ \frac{1}{a_0}\right]^T \tag{11}$$

For any $\mathbf{r}_j \in R^{(i)}$ and any $\mathbf{r}_l \in R^{(i+1)}$, we have $H_i(\mathbf{r}_j) = -1 < 0$ and $H_i(\mathbf{r}_l) = 1 > 0$.

The MMSE solution $\mathbf{w}_{\text{MMSE}}$ is another example of separating hyperplane. The non-optimal nature of the MMSE solution is clearly illustrated in the asymptotic case of large signal to noise ratio (SNR) [1]: $\mathbf{w}_{\text{MMSE}} \to \hat{\mathbf{w}}$ as SNR$\to \infty$.

The SVM solution

We will use R_- and R_+ to represent $R^{(M/2)}$ and $R^{(M/2+1)}$. The minimum distance from the nearest point in $R_\pm$ to a separating hyperplane, $\mathbf{w}^T \mathbf{r} = 0$, is given by:

$$\rho(\mathbf{w}) = \min_{\mathbf{r}_i \in R_+} \frac{|\mathbf{w}^T \mathbf{r}_i|}{\|\mathbf{w}\|} + \min_{\mathbf{r}_j \in R_-} \frac{|\mathbf{w}^T \mathbf{r}_j|}{\|\mathbf{w}\|} \tag{12}$$

which is referred to as the margin. The SVM design finds the hyperplane that maximizes this margin. Since the weight vector $\mathbf{w}$ of the hyperplane $\mathbf{w}^T \mathbf{r} = 0$ is linear dependent, it is appropriate to consider a canonical hyperplane [2] where $\mathbf{w}$ is constrained by

$$\min_{\mathbf{r}_i \in R_\pm} |\mathbf{w}^T \mathbf{r}_i| = 1 \tag{13}$$

Define the integer set $I_{R_\pm} \triangleq \{i : \mathbf{r}_i \in R_\pm\}$ and the class indicator $y_i = \pm 1$, $\forall \mathbf{r}_i \in R_\pm$. The maximization of the margin (12) with the constraint (13) using the classical Lagrangian theory [4] gives rise to the optimal separating hyperplane:

$$\mathbf{w}_{\text{SVM}} = \sum_{i \in I_{R_\pm}} \bar{g}_i\, y_i\, \mathbf{r}_i \tag{14}$$

where

$$\bar{\mathbf{g}} = \arg\min_{\mathbf{g}} \frac{1}{2} \sum_{i \in I_{R_\pm}} \sum_{j \in I_{R_\pm}} g_i g_j\, y_i y_j\, \mathbf{r}_i^T \mathbf{r}_j - \sum_{i \in I_{R_\pm}} g_i \tag{15}$$

$$g_i \geq 0,\ \forall i \in I_{R_\pm} \tag{16}$$

The optimization problem (15) with (16) is a convex quadratic programming, whose solution $\bar{\mathbf{g}}$ can be computed efficiently. As g_i are the Lagrange multipliers of the primal problem, from the Kuhn-Tucker conditions [4]

$$\bar{g}_i \left(y_i \mathbf{w}_{\mathrm{SVM}}^T \mathbf{r}_i - 1 \right) = 0 \tag{17}$$

only those $\mathbf{r}_i$, which satisfy $y_i \mathbf{w}_{\mathrm{SVM}}^T \mathbf{r}_i = 1$, will have non-zero Lagrange multipliers. These points are the support vectors (SVs) [2]. All the SVs lie on the margin and the number of SVs can be very small. The hyperplane $\mathbf{w}_{\mathrm{SVM}}$ is uniquely determined by the set of SVs.

From lemma 1, it can be seen that the optimal hyperplane for separating $R^{(i)}$ and $R^{(i+1)}$ is

$$\bar{H}_i(\mathbf{r}) = \mathbf{w}_{\mathrm{SVM}}^T \mathbf{r} + c_i = 0 \tag{18}$$

with the bias

$$c_i = 2 \left(\frac{M}{2} - i \right) \mathbf{w}_{\mathrm{SVM}}^T \mathbf{a}_{\mathrm{rev}} \tag{19}$$

We comment that $\mathbf{w}_{\mathrm{SVM}}$ is independent of the noise variance and, in the asymptotic case of large SNR, it becomes the MSER solution $\mathbf{w}_{\mathrm{MSER}}$. In general, $\mathbf{w}_{\mathrm{SVM}}$ will not be identical to $\mathbf{w}_{\mathrm{MSER}}$, but the difference is practically negligible for useful SNR conditions.

A simulation example

The example had a 3-tap CIR of $\mathbf{a} = [-0.35\ 0.8\ 0.35]^T$ with 4-PAM symbols. The DFE structure was defined by $d = 2$, $m = 3$ and $n = 2$. The subsets $R_\pm$ have 32 states, and the SVM solution $\mathbf{w}_{\mathrm{SVM}} = [0.5543\ 1.2669\ \ -1.0699]^T$ is determined by the 8 SVs. The MMSE solution for SNR=23dB is $\mathbf{w}_{\mathrm{MMSE}} = [0.2227\ 0.5902\ \ -1.2405]^T$. Fig. 1 depicts the SER curves of the MMSE and SVM DFEs with detected symbols being fed back. For this example, the SERs of the MSER DFE, not included here, are practically indistinguishable from those of the SVM DFE.

Conclusions

We have extended the SVM technique to design the linear-combiner DFE for M-PAM signals. Although the SVM DFE only achieves the MSER performance asymptotically, in practice, the performance of the two DFEs are almost identical. The SVM DFE generally outperforms the MMSE DFE, but it does not have a simple adaptive implementation and can only be implemented indirectly by first estimating a CIR and then designing the equalizer weight vector. This design process however is a simple quadratic programming.

References

[1] S. Chen and B. Mulgrew, "The minimum-SER linear-combiner decision feedback equalizer," *IEE Proc. Communications*, Vol.143, No.6, pp.347–353, 1999.

[2] V. Vapnik, *The Nature of Statistical Learning Theory.* New York: Springer-Verlag, 1995.

[3] S. Chen, S. Gunn and C.J. Harris, "Decision feedback equalizer design using support vector machines," *IEE Proc. Vision, Image and Signal Processing*, to appear, 2000.

[4] M. Minoux, *Mathematical Programming: Theory and Algorithms.* Chichester: John Wiley and Sons, 1986.

S. Chen, A.K. Samingan and L. Hanzo (*Department of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K.*)

Tel./Fax: +44 (0)23 8095 6660/4508; E-mail: sqc@ecs.soton.ac.uk.

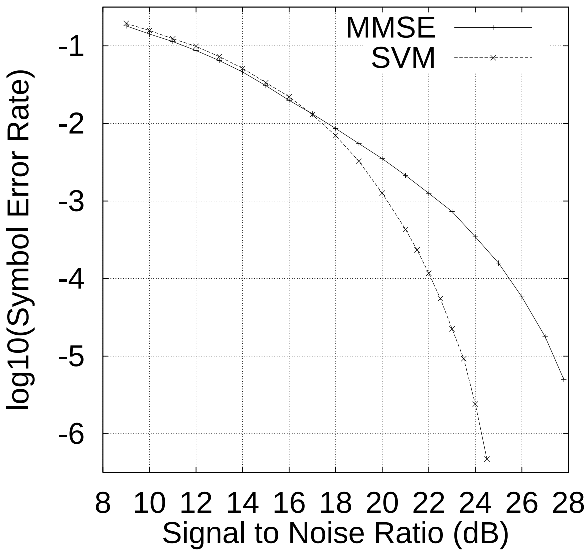

Figure 1: Performance comparison for the CIR of $\mathbf{a} = [-0.35\ \ 0.8\ \ 0.35]^T$ with 4-PAM symbols.